ELSEVIER

Tangled webs: Tracing the connections between genes and cognition

Simon E. Fisher [*]

Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK

Abstract

The rise of molecular genetics is having a pervasive influence in a wide variety of fields, including research into neurodevelopmental disorders like dyslexia, speech and language impairments, and autism. There are many studies underway which are attempting to determine the roles of genetic factors in the aetiology of these disorders. Beyond the obvious implications for diagnosis, treatment and understanding, success in these efforts promises to shed light on the links between genes and aspects of cognition and behaviour. However, the deceptive simplicity of finding correlations between genetic and phenotypic variation has led to a common misconception that there exist straightforward linear relationships between specific genes and particular behavioural and/or cognitive outputs. The problem is exacerbated by the adoption of an abstract view of the nature of the gene, without consideration of molecular, developmental or ontogenetic frameworks. To illustrate the limitations of this perspective, I select two cases from recent research into the genetic underpinnings of neurodevelopmental disorders. First, I discuss the proposal that dyslexia can be dissected into distinct components specified by different genes. Second, I review the story of the *FOXP2* gene and its role in human speech and language. In both cases, adoption of an abstract concept of the gene can lead to erroneous conclusions, which are incompatible with current knowledge of molecular and developmental systems. Genes do not specify behaviours or cognitive processes; they make regulatory factors, signalling molecules, receptors, enzymes, and so on, that interact in highly complex networks, modulated by environmental influences, in order to build and maintain the brain. I propose that it is necessary for us to fully embrace the complexity of biological systems, if we are ever to untangle the webs that link genes to cognition.
© 2006 Elsevier B.V. All rights reserved.

[*] Tel.: +44 1865 287647; fax: +44 1865 287650.
   E-mail address: simon.fisher@well.ox.ac.uk

1. Introduction

In recent years we have witnessed some extraordinary advances in the field of molecular genetics (Collins, Green, Guttmacher, & Guyer, 2003). Since this research area has the potential to impact on a diverse range of other disciplines within the natural sciences, such advances are having a pervasive influence on multiple fields of biology. Human cognition is no exception; the past decade has seen a dramatic increase in the number and scale of research programmes exploiting molecular approaches for studying the human brain. There are now many laboratories worldwide investigating the role of genetic factors in developmental learning disorders (Fisher & DeFries, 2002; Fisher, Lai, & Monaco, 2003; Folstein & Rosen-Sheidley, 2001; Grigorenko, 2001) using the same molecular techniques that are being applied to common human traits such as heart disease, hypertension, diabetes, asthma, arthritis and so on (Botstein & Risch, 2003). Biologists are now in possession of an exciting and powerful new toolkit for tackling their favourite research questions. However, as is often the case, there is a catch. Although the emerging molecular methods are deceptively easy to wield, the answers they present are far from straightforward. Making sense of them requires a novel mindset, one which is rooted in complexity and unfamiliar to many. In the present article I begin by outlining the basic concepts that underlie molecular genetic analyses of developmental disorders. I then go on to argue that the common misconception of the gene as an abstract entity, able to elusively control aspects of cognition, may impede attempts to connect genes with cognition. Finally, I illustrate the key issues by considering a selection of pertinent examples from recent studies of developmental dyslexia and speech and language disorders. This is not intended as a general review of the field; I specifically focus on cases that best demonstrate the pitfalls of adopting an abstract view of the gene, and that highlight the importance of integrating data from multiple levels of analyses. I propose that we need to fully embrace the inherent molecular complexity of developmental systems, if we want to further our understanding of genetic influences on human cognition.

2. Back to basics

As I illustrate elsewhere in this article, it is quite possible to carry out genetic analyses with only an abstract perception of the nature of a gene. The relevant methods are easily transposable to virtually any trait of interest (assuming that such a trait has at least some heritable basis). However, a proper appreciation of the significance of genetic findings must depend on a solid foundation in basic molecular concepts, and it is worth briefly revisiting these before proceeding further. Given the blaze of

publicity surrounding the Human Genome Project, the vast majority of people are aware that (i) genetic information is encoded by the sequence of nucleotides in a molecule known as DNA, (ii) the genome represents the entire DNA sequence of a particular organism; the full set of "instructions" for building and maintaining that organism, (iii) almost every cell of a multi-cellular organism has its own copy of the genome to refer to, and (iv) DNA molecules are tightly packaged (with proteins and other molecules) into bodies known as chromosomes. Many people will also know that the human genome, recently decoded, contains over 3 billion nucleotide letters. But there are common gaps in understanding. For example, how does the concept of "the gene" fit into this framework of DNA, chromosomes and genomes? How can the linear code contained in a string of molecular letters specify the construction and functions of a three-dimensional organism (in fact four-dimensional, since we need to take into account temporal aspects)? If the whole human genome has indeed been decoded and all the human genes have been found, then why are scientists still "mapping" the genes?

There are also a number of frequent misunderstandings of genetics, which persist from a time when we did not have a molecular perspective, such as the idea that there might literally be a "gene *for* dyslexia". This kind of term is a potentially misleading piece of geneticists' jargon, originating from the early days of the field, in which a gene is described purely in terms of the overall dysfunction associated with its loss or mutation. Our ability to characterise genes has since moved on dramatically, but the "gene for X" phraseology continues to be very extensively used both in the literature and the media, mainly due to its effectiveness in communicating the importance of genetic influences. A wider but related issue is whether genes may exist whose job it is to directly specify particular cognitive abilities (e.g. a "gene for grammar", or a "gene for phoneme awareness"). I will return to these points at the end of this article, since they are central to common misconceptions regarding the role of genes in cognition.

So, what is a gene? Answering this question is far from trivial, but a useful operational definition might be "a stretch of DNA whose linear sequence of nucleotides encodes the linear sequence of amino acids in a specific protein". The amino acid sequence of a protein determines the way that it folds, and hence its three-dimensional structure. The function of the resulting protein is tightly dependent on this three-dimensional structure. As a consequence of variability in the combinations of amino acids that can be strung together into polypeptides, genes are able to encode proteins with a vast range of differing functions. These include enzymes for catalyzing biochemical reactions, proteins that are integral to the structure or motility of a cell, molecules that are sent out as signals to other cells, receptors that receive those signals, channels that allow ions to pass in and/or out of a cell and many more, together forming components of the "molecular machines of life" (Hood & Galas, 2003). The genome is pivotal in that it contains all the instructions for assembling the particular arrays of molecular machines that are characteristic for each species. More importantly, the genome also encodes the information that controls the way in which each gene is expressed (turned on and off) in time and space, linking them into complex regulatory networks. Such networks comprise two major constituents; (i) stretches

of noncoding DNA that control the quantity of the encoded protein that any gene makes, (ii) regulatory proteins that bind to these control regions and modulate their activity. The combined activity of control regions and regulatory factors is at the heart of the development and physiology that defines an organism (Hood & Galas, 2003). Thus, it is important to realise that the appearance and biology of a mature organism is the result of a complex series of ontogenetic events unfolding over time, moderated by environmental and stochastic influences. Genomes are much more like knitting patterns or recipes than blueprints (although even the former are poor analogies for the peculiarities of the genome). The majority of the sequence of the human genome has been determined, but it still harbours many surprises; for example it has recently become clear that, in addition to the well-understood protein-coding genes, genomes also encode a host of RNA transcripts that do not direct the synthesis of proteins, but nevertheless play key roles in regulating cell behaviour (e.g. Kuwabara, Hsieh, Nakashima, Taira, & Gage, 2004).

The "Human Genome Sequence" (Wolfsberg, McEntyre, & Schuler, 2001) is a generalised description somewhat akin to anatomical annotations of the structure of the organs and skeleton of the human body. On the basis of sequencing data, the human genome contains 24–30,000 different protein-encoding genes. (Remember that the definition of a gene is a complex issue, and it can sometimes be hard to predict where a gene begins and ends, hence the continuing uncertainty over the exact number of genes in the genome.) The vast majority of people have essentially the same *set* of protein-encoding genes and regulatory regions (with a few exceptions, which are not relevant to the present discussion). However, due to random mutations, the actual nucleotide sequences tend to vary between individuals of the same species in a given population. As a consequence different people may have distinctive versions of any particular gene, known as "alleles" or "allelic variants". Although many allelic variants are functionally silent, meaning that they do not alter the resulting protein sequence, or the way that the protein is regulated, some variants lead to functional changes. For example, a simple nucleotide change may alter the type of amino acid at a defined point within the encoded protein, modifying both its structure and function. Tiny sequence changes at the nucleotide level can even result in a complete absence of protein function, for example by causing the amino acid chain to be prematurely terminated. Allelic variants with such dramatic consequences for protein function tend to be present at low frequency in the human population as a consequence of negative selection and are responsible for classical single gene disorders like Duchenne muscular dystrophy. However, alleles with more subtle effects are likely to be present at higher frequency, and could contribute to variability in common traits (Reich & Lander, 2001).

3. The human geneticist's toolkit

A central aim of modern human genetics is to determine the causal links between genotype (the genetic makeup of an individual in terms of allelic variants) and phenotype (the appearance of an individual in terms of a particular characteristic, be it

physical, biochemical, physiological or neurodevelopmental) (Botstein & Risch, 2003). In pursuing this aim, geneticists seek to address a number of separate, but related, issues (Collins et al., 2003). At a clinical level, such efforts hope to identify the molecular basis of diseases and developmental disorders, and success in this area can have obvious ramifications for diagnosis and intervention (Bell, 2003). Linking genotype to phenotype is also essential if we are to understand how natural genetic variation contributes to trait variability in normal populations (Freimer & Sabatti, 2003). At a more fundamental level, these kinds of studies promise to help uncover the roles of different genes in aspects of human biology, shedding light on embryonic development, metabolism, immunity, tissue-specific functions and so on (Collins et al., 2003). Thus, it can be argued that identification of genetic variants predisposing to syndromes like dyslexia, language impairment and autism will not only enable us to improve diagnosis and treatment of these disorders, but could also give new insights into normal cognitive processes and increase our understanding of molecular pathways that impact on brain function (Fisher & Marcus, 2006; Marcus & Fisher, 2003).

When attempting to find links between genotypes and phenotypes, researchers have at their disposal a variety of methods and techniques, which are generally applicable to any trait or gene of interest. In particular, ''positional'' strategies have proved to be lucrative for a wide range of situations, since they can be successful even in the total absence of any knowledge regarding the biological underpinnings of a trait (Botstein & Risch, 2003). A typical study of this kind involves analyses of the inheritance patterns of different chromosomal regions in families affected by the trait being studied. Remember that every human being inherits two copies of each of the nonsex chromosomes, one originating from the father (paternal) and one originating from the mother (maternal). These copies will usually contain an identical array of genes, but they may differ with respect to the specific allelic variants at each gene. The choice of chromosomal copy that is passed on to the next generation is essentially random, and each chromosome assorts independently of the others. In addition, homologous (corresponding) sections of maternal and paternal chromosome copies are sometimes exchanged during generation of germ cells (oocytes/sperm) by a process called crossing over. The occurrence of crossing over events is generally stochastic, although they can cluster in ''hotspots'', and the overall result is that corresponding sections of chromosomes are shuffled around from generation to generation. This process of genetic recombination means that the genomic make-up that is inherited by an individual can be viewed as a patchwork of maternal and paternal alleles from previous generations (Pääbo, 2003). It is the patchwork nature of genomic transmission that allows geneticists to map allelic variants influencing a trait of interest to particular chromosomal intervals.

In practice this can be achieved by (i) identifying families containing individuals affected with a trait of interest, (ii) determining the inheritance patterns of each section of chromosome within each family, and (iii) asking whether the inheritance of any chromosomal interval is correlated (or ''linked'') with the inheritance of the trait. Geneticists refer to this process as ''linkage analysis''. Positive correlations may result from a real biological effect (i.e. allelic variation of a gene mapping within

the chromosomal interval could be influencing the trait) but might also occur by chance. Thus, it is necessary to perform statistical tests to assess how likely it is that a linkage result is a chance finding (Lander & Kruglyak, 1995). Typically linkage studies exploit naturally occurring variations in DNA sequence at a series of markers along a chromosome. These markers do not need to be within genes or have any biological function; they are simply used as a means to track inheritance patterns, a way of mapping out the patchwork of chromosomal chunks that get passed through a family. A strength of linkage-based approaches is that by screening only a small number of markers one can obtain a clear picture of the inheritance patterns within a family; for example using only 400 markers it is possible to rapidly scan the entire human genome (all 3 billion letters) for a genetic effect on a trait of interest. The concomitant weakness is that, by themselves, these techniques are unable to exactly pinpoint the position of the putative genetic change that influences the trait, but instead provide the first clue to its location; regions implicated by linkage may contain tens or even hundreds of genes, any of which might harbour the causative variant. Nevertheless linkage analysis can often narrow the focus of the search quite substantially (given that the full human genome may contain as many as 30,000 protein-encoding genes). This approach has therefore played an extremely important role in helping to identify the genes implicated in a large number of simple genetic disorders, and is beginning to make its mark in more complex traits (Korstanje & Paigen, 2002).

Classical linkage analysis in humans relies on dichotomous classification of whether an individual is affected or unaffected (referred to as "affection status"). While this has proved a powerful tool under certain circumstances, there are often difficulties with defining disorder in complex behavioural or cognitive syndromes (Fisher & DeFries, 2002; Fisher et al., 2003; Folstein & Rosen-Sheidley, 2001; Grigorenko, 2001). Moreover, categorical diagnoses are unable to speak to genetic influences on variability in the normal range. Many recent studies of complex human traits have been able to move away from categorical definitions altogether, by using quantitative-trait-locus (QTL) methods (Glazier, Nadeau, & Aitman, 2002). These are similar to methods based on affection status, in that they assess the significance of phenotype-genotype correlations for the chromosomal region under investigation. However, they directly incorporate quantitative data (e.g. test scores obtained with measures of cognitive ability) in the statistical analysis, rather than collapsing the phenotype into the dichotomous status of affected versus unaffected. The choice of whether to use categorical definitions or quantitative traits depends to some extent on the particular study design and the questions that are being asked; each approach has its own strengths and weaknesses (see Fisher & DeFries, 2002; for more discussion). For neurodevelopmental disorders, genome-wide scans for linkage have been effective in identifying chromosomal intervals that might be implicated in dyslexia (de Kovel et al., 2004; Fagerheim et al., 1999; Fisher et al., 2002b; Kaminen et al., 2003; Nopola-Hemmi et al., 2001), specific language impairment (Bartlett et al., 2002; SLI Consortium, 2002) and autism (numerous scans including Buxbaum et al., 2001; CLSA, 2001; IMGSAC, 1998, 2001; Philippe et al., 1999; Shao et al., 2002; Yonan et al., 2003). These studies have used a mixture of different approaches, some employing categorical definitions, others relying on QTL-based methods. It is

worth a note of general caution here regarding interpretation of linkage findings in studies of complex traits. Independent replication is essential to support the robustness of any proposed linkage, and there have been many cases where findings could not subsequently be supported in separate datasets.

Higher resolution mapping of genetic effects is feasible using methods based on ''association''. While linkage assesses marker-trait correlations within families, association evaluates whether there is a correlation between a particular allelic variant and a trait at the population level (Cardon & Bell, 2001). One of the simplest study designs for detecting association is to compare allele frequencies at a genetic locus in affected cases with those in unaffected controls. A higher frequency of an allele among the cases might indicate that this variant acts as a risk factor for the disorder (conversely a lower frequency could suggest a protective effect). In a similar manner to linkage, these methods have been extended for use with quantitative traits. As with linkage, there is the danger that an allele-trait association arises due to chance, so robust statistical analysis is again of the essence (Cardon & Bell, 2001). (Spurious associations may also result from factors like population stratification, which will not be discussed here.) In addition, a significant association does not necessarily indicate that the allele in question is the aetiologically important variant; it may be due to a phenomenon known as ''linkage disequilibrium'' in which neighbouring allelic variants have travelled together during the recent evolutionary history of the population. Association-based methods are very powerful for detecting small genetic effects, but are only sensitive if the marker under investigation maps close to the site of the aetiological variant in the genome. In other words, if taking a purely positional approach to systematically scan the entire genome via association, one would have to screen an unfeasibly large number of markers to be confident of finding a real effect. (Note, however, that as technology improves this is becoming less of a hurdle.) Many researchers in the field of complex traits are currently using a combined strategy, in which the entire genome is screened via linkage, and then the implicated regions are the focus of intensive association-based efforts. For example, chromosomal intervals implicated by linkage studies of developmental dyslexia are now being closely examined via high-throughput association analyses (e.g. Cope et al., 2005a; Deffenbacher et al., 2004; Francks et al., 2002, 2004).

The study of gross chromosomal abnormalities represents another invaluable method for making connections between genomic regions and particular traits. In a small proportion of individuals, disorder is caused by disruption to genes as a consequence of deletions, inversions, duplications or translocations of sections of chromosomes. Well-characterised examples of the effects of such abnormalities on brain function include Down's syndrome (Roizen & Patterson, 2003) and Williams syndrome (Morris & Mervis, 2000), but the total impact of sporadic chromosomal rearrangements on cognitive deficits is unknown. Even when rearrangements are not detected under the microscope, subtle abnormalities (e.g. submicroscopic deletions) may be present; for example rearrangements of chromosome ends (telomeres) account for a significant proportion of cases of idiopathic mental retardation (Flint & Knight, 2003). Investigation of chromosomal abnormalities can be particularly powerful if they coincide with intervals already implicated by linkage in mapping

studies; however, the identification of such cases depends to a large extent on seren-dipity. In addition, many sporadic chromosomal rearrangements are benign, so that their occurrence in affected subject is sometimes coincidental rather than causal. Nevertheless, this is an essential branch of genetics that is playing a major role in our attempts to connect genes to cognition; for example studies of chromosomal translocations provided the basis for the identification of the *DYX1C1* gene as a can-didate for involvement in dyslexia (Taipale et al., 2003) and were also important in the discovery of the role of *FOXP2* in speech and language disorder (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001).

Attempts to connect specific phenotypes with distinct genotypes can also be driven by hypotheses regarding the biological basis of a trait. In these cases, theories regarding the aetiology of a disorder suggest particular processes or pathways, impli-cating a gene or set of genes which can be considered as candidates. Direct associa-tion analysis or mutation screening of a candidate gene may yield evidence supporting involvement of that gene, and may also strengthen the case for the aeti-ological hypothesis. This kind of approach has been extensively adopted for investi-gating Attention Deficit/Hyperactivity Disorder (ADHD), with candidate genes being selected based on knowledge of effective drug treatments (DiMaio, Grizenko, & Joober, 2003), although complementary positional-based strategies are also begin-ning to be used (Arcos-Burgos et al., 2004; Bakker et al., 2003; Fisher et al., 2002a; Ogdie et al., 2003). For many neurodevelopmental phenotypes of interest, including dyslexia, specific language impairment (SLI) and autism, we have highly limited understanding of the likely molecular underpinnings. Thus, in these cases, pure can-didate gene approaches (in absence of any positional information from linkage, asso-ciation or chromosomal abnormality-based studies) have not yet been successful for genetic dissection, while combined strategies appear to be more productive (e.g. see dyslexia studies of Cope et al., 2005a; Francks et al., 2004; Taipale et al., 2003).

4. Discarding the abstract gene: Acknowledging complexity and context

Equipped with the above tools of contemporary human genetics, researchers are starting to make headway in the hunt for genetic variants that contribute to common neurodevelopmental disorders of unknown aetiology. The present article does not seek to provide yet another overview of all the molecular genetic results obtained from these studies, since comprehensive accounts of progress in the fields of autism (Folstein & Rosen-Sheidley, 2001; Wassink, Brzustowicz, Bartlett, & Szatmari, 2004), dyslexia (Fisher & DeFries, 2002; Fisher & Francks, 2006; Grigorenko, 2001) and speech and language disorders (Fisher et al., 2003) have been published elsewhere. Moreover, I focus here on what we can learn about genes and cognition from the point of view of human neurodevelopment. Investigations of neurodegen-erative disease (such as the well characterised role of the *APOE* gene in dementia), or genetic influences on normal cognitive function (e.g. potential effects of *BDNF* gene variants on memory processing) will not be discussed. For an insightful review of the latter, readers are referred to Goldberg and Weinberger (2004).

So, rather than providing an additional general review, I propose here to highlight a number of key issues that have emerged from recent studies of neurodevelopmental genetics, in particular those that illustrate the need to replace views based on ''the abstract gene'' with more complex and realistic paradigms. The first point is that while the majority of common neurodevelopmental disorders are known to have a significant heritable component (based on twin and adoption studies), the genetic contributions are largely multifactorial in nature. That is, in most cases the susceptibility to these disorders will involve variants in multiple genes each with a small to moderate effect, modulated by the influences of environmental factors. Different clusters of genes may be implicated in different affected subjects (genetic heterogeneity) and increased risk might depend on interactions between several genes. Gene–environment interactions may also be critical for the manifestation of disorder. The presence of this genetic complexity, which can obscure the connections between genotypes and phenotypes, was initially suspected on the basis of epidemiological studies, and has been strongly confirmed by molecular mapping studies (Fisher & DeFries, 2002; Fisher et al., 2003; Folstein & Rosen-Sheidley, 2001; Grigorenko, 2001). There are some notable rare exceptions; families which display apparently simple monogenic (single gene) inheritance of neurodevelopmental impairment (e.g. Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998) which I will discuss later.

The second issue worth highlighting is that while considerable progress has been made in recent years in *mapping* genetic influences on common neurodevelopmental disorders, the identification of specific alleles that confer susceptibility remains problematic. Although other approaches are now coming to the fore, the majority of investigations thus far have relied on linkage analyses. For example, linkage-based studies of dyslexia have mainly implicated regions on chromosomes 1, 2, 3, 6, 15 and 18 (reviewed by Fisher & DeFries, 2002 & Grigorenko, 2001), while genome screens in SLI point to potential loci on 13, 16 and 19 (Bartlett et al., 2002; SLI Consortium, 2002). A few of the dyslexia linkages have shown consistency in independent studies, to an extent that is unusual for a complex trait, suggesting that they represent real effects. Most notably, linkage to a region on the short arm of chromosome 6 has been supported in multiple datasets and analyses (e.g. Cardon et al., 1994; Fisher et al., 1999; Gayán et al., 1999; Grigorenko et al., 1997; Kaplan et al., 2002; Turic et al., 2003), although not every study has been able to replicate (Field & Kaplan, 1998; Schulte-Körne et al., 1998). Nevertheless, unambiguous cases of replicated linkages remain the exception. Consider the two independent SLI genome screens of Bartlett et al. (2002) and the SLI Consortium (2002); strong linkage peaks were uncovered in each scan (chromosome 13 in the former, and chromosomes 16 and 19 in the latter) but there was virtually no overlap in findings from these two large-scale studies. Moreover, despite overall advances in mapping, no study has yet demonstrated a clear replicable effect of one particular functional variant on a common form of neurodevelopmental disorder. The situation is promising for the case of the dyslexia-related locus on chromosome 6, where independent association analyses have converged on a small interval containing only a few genes (including *KIAA0319* and *DCDC2*), but even then there is disagreement over which of these genes influenc-

es risk, and the functionally significant variants remain to be defined (Cope et al., 2005a; Deffenbacher et al., 2004; Fisher & Francks, 2006; Francks et al., 2004; Meng et al., 2005). A handful of genes have been directly implicated in rare cases of neurodevelopmental disorder, including *FOXP2* in speech and language impairment (Lai et al., 2001), neuroligins *NLG3* and *NLG4* in autism (Jamain et al., 2003) and *DYX1C1* and *ROBO1* in dyslexia (Taipale et al., 2003; Hannula-Jouppi et al., 2005). However, coding variants of *FOXP2* have not been found in typical SLI (Newbury, Bonora, Lamb, Fisher, & Lai, 2002; O'Brien, Zhang, Nishimura, Tomblin, & Murray, 2003), and the relevance of variants in *NLG3*, *NLG4*, *DYX1C1* and *ROBO1* for common disorders remains open to question (Cope et al., 2005b; Gauthier et al., 2004; Hannula-Jouppi et al., 2005; Marino et al., 2005; Scerri et al., 2004; Vincent et al., 2004; Wigg et al., 2004). More discussion regarding the relationship between specific candidate genes (*DYX1C1*, *KIAA0319*, *DCDC2* and *ROBO1*) and dyslexia susceptibility can be found in Fisher and Francks (2006).

My third point expands on an aspect of genetic studies that I alluded to in previous sections of this article. Positional approaches have the advantage that they can be carried out in total absence of any knowledge regarding aetiological mechanisms; genotype can be connected to phenotype with no need for reference to biological or developmental context. As our knowledge of the genome increases and genotyping costs plummet, genetic investigations of human brain-related phenotypes are becoming more and more widespread. The only caveat appears to be the need to demonstrate with robust statistics that any correlations that are detected are unlikely to be due to chance. Unfortunately, the ability to undertake genetic analyses while employing only the most basic abstract concept of "the gene", and without any understanding of molecular pathways, has become both a blessing and a curse, particularly in studies of the brain. The apparent ease of correlating genotype with phenotype without reference to molecular/developmental mechanisms promotes an erroneous impression of neurogenetics; one in which individual genes are able to mysteriously control specific behaviours or cognitive abilities, leading to talk of "language genes", "smart genes", "gay genes", "aggressive genes" and so on. It is indisputable that variations of gene sequence can contribute to variability in cognitive abilities and personality traits (sometimes in a dramatic manner) and that apparently straightforward genotype-phenotype correlations can sometimes emerge in our datasets. But the simplicity of these relationships is merely an illusion; genes do not (and indeed can not) *specify* particular behavioural outputs or cognitive processes, except in the most indirect way. As highlighted by Inoue and Lupski (2003), assumptions of simple linear relations between genes and cognitive/behavioural phenotypes have impeded progress in the field, and fuel hypotheses that must ultimately be untenable. The gross activities of the human brain are the products of a complex interplay between factors at multiple levels; be they genetic, cellular, developmental, anatomical, or environmental, and the routes linking genes to cognition will inevitably be tortuous (Marcus, 2003). It is worth noting that this is likely to apply even to supposedly simple monogenic disorders of brain development, as I illustrate below with the example of the *FOXP2* gene. This is not to imply that any attempts to disentangle links between genes and cognition are a waste of time. However, we ignore at our

peril the existence of molecular and ontogenetic complexity and the importance of developmental context. Grant (2003) has similarly argued that the gap between genes and cognition can only be bridged by a thorough systems biology account of brain development and function.

Even pure candidate gene approaches can be victims of the ''abstract gene'' perspective. In many cases, when researchers find statistical evidence to support association between a particular variant of a gene and a common trait, it is erroneously assumed on the basis of this that the variant is likely to be causative and that there is a simple pathway connecting gene to trait. In studies of common human behavioural/cognitive traits, it is rare to see experiments demonstrating that putative risk variants actually do affect the function or expression of candidate genes, although investigators are beginning to remedy the situation (e.g. Kakiuchi et al., 2003). This brings me on to my final issue, which may be referred to as the ''Are we there yet?'' phenomenon (Page, George, Go, Page, & Allison, 2003). In single-gene disorders the aetiological variants often have obvious and gross consequences for gene function; for example via protein truncation or alteration of a critical functional domain. In addition, the correlation between inheritance of these mutations and development of the disorder in these cases usually approaches 100%. Under these circumstances a geneticist can be confident that the cause of the disorder has indeed been pinpointed. (Note, however, that complementary functional data are invariably sought to provide the final proof.) For traits with a complex genetic basis, the connections between gene variation and phenotype are much more difficult to pin down; the risk variant might be commonly found amongst unaffected individuals, absent from many affected individuals, and the consequences for expression or function of the relevant gene could be subtle. Several neighbouring allelic variants may be in linkage disequilibrium, each yielding significant statistical evidence for association with the trait of interest, so that it can be highly challenging to determine the identity of the true functional variant (or indeed if the entire association is due to a false positive result). This predicament is at the core of all attempts to genetically dissect complex traits (Page et al., 2003). There is a large gulf between finding statistical evidence for a genotype-phenotype correlation and demonstrating a convincing causal relationship, and appreciation of the distinction is essential for those trying to interpret studies in this field (Fisher & Francks, 2006).

5. Dissecting disorders: Cautionary tales from studies of dyslexia

When searching for correlations between neurodevelopmental traits and genotypes, the approach to phenotype definition is critical. In investigations of disorder, a dichotomous classification of affected versus unaffected provides the most simple basis for detecting linkage (identifying chromosomal intervals that are commonly inherited by multiple affected members of a family) or association (for example in comparisons of allele frequencies in cases versus controls). However, there are often difficulties associated with assigning affection status in complex traits (Lander & Schork, 1994). For disorders such as dyslexia, autism or SLI there can be consider-

able variability in the severity of impairment and the profile of cognitive abilities amongst different affected individuals (for reviews see Fisher & DeFries, 2002; Fisher et al., 2003; Folstein & Rosen-Sheidley, 2001; Grigorenko, 2001). Moreover, the same individual may show changes in phenotypic profile at distinct points of his or her development. Attempts to subdivide neurodevelopmental disorders into separable subtypes with clearly distinctive aetiologies have met with only limited success. A number of complementary strategies have been employed to handle issues of phenotypic complexity for genetic analyses, and studies of dyslexia provide some apt illustrations.

One approach is to decompose the phenotypic profile that is associated with the disorder into different features, initially referred to in the literature as "components" (Grigorenko et al., 1997). The nature of so-called components is essentially driven by hypotheses regarding potential cognitive bases of disorder. The most overt characteristics of dyslexia are impairments in reading and/or spelling that are not attributable to general intellectual delay or inadequate environmental stimulation (Habib, 2000). However, it has been robustly shown that most affected individuals have significant linguistic deficits even when performing tasks that do not involve any reading or writing (Snowling, 2001). There has thus been much focus in recent years on tests that tap into hypothesized phonological processing pathways which appear to be impaired in subjects with dyslexia. These include phoneme awareness, defined as the ability to explicitly manipulate the separate speech units that make up a word. In addition, many individuals with dyslexia manifest difficulties when rapidly naming sequences of objects, colours, shapes and so on (Wolf & Bowers, 2000). Hypothetical phenotypic components have also been derived from models describing the cognitive underpinnings of normal reading. A relevant example is the dual-route model, which proposes that the mature reader has available two alternative routes for reading a visually presented word. One (phonological decoding) involves the separation of the written word into constituent grapheme units and conversion of these to appropriate phonemes, while the other (orthographic coding) requires recognition of the entire word and retrieval of the appropriate phonological representation from a mental lexicon (Castles & Coltheart, 1993). It has been argued that decoding of pronounceable nonsense words (nonwords) should primarily tap the putative phonological route, while accurate reading of irregular words (those that violate standard grapheme–phoneme conversion rules, like "colonel" or "yacht") will predominantly involve orthographic processing. Questions regarding the validity of dual-route models of reading will not be discussed here. The key point is that theories of normal and abnormal reading processes have led geneticists to explore phenotypes such as phoneme awareness, phonological decoding, orthographic coding and rapid automised naming in families affected with dyslexia. Note that although the phenotypes used in these studies have been termed "components" of dyslexia, they do not represent fully independent constituents of the cognitive profile; Grigorenko et al. (1997) described them as "more precise defining attributes, partly overlapping but partly distinct". Indeed, there tend to be moderate to high correlations between the different language- and reading-related measures that are used as indices of putative components; for example nonword reading and irregular word reading, often

employed to assess phonological decoding and orthographic processing abilities respectively, show correlations of 0.57–0.61 in one large set of families (Marlow et al., 2001). A small proportion of individuals appear to show dissociation between nonword reading and irregular word reading ability, leading to suggestions of phonological- and surface-dyslexia subtypes (Castles & Coltheart, 1993), but in most subjects with dyslexia, deficits are observed for both nonword and irregular word reading.

The strategy exemplified by the work of Grigorenko et al. (1997) involves multiple sets of genetic analyses of the same families under a series of different diagnostic schemes, each tapping one of the above-mentioned hypothetical components. Thus, one set of analyses might assess phenotype–genotype correlations for an affection status based on impaired phoneme awareness, another might look at the phenotype of phonological decoding, while a third might examine deficits in rapid automised naming. It is worth highlighting that this is somewhat different from studying separate subtypes, since the components are partly overlapping e.g. some individuals within a family might be considered as affected only for phoneme awareness, others might be affected for both phoneme awareness and phonological decoding, and so on. Most importantly there are usually members of families who are not considered to be dyslexic, but do show deficits on one or more of the putative component phenotypes (Grigorenko et al., 1997). Thus, Grigorenko and colleagues were able to address phenotypic complexity while continuing to exploit the power and simplicity of categorical definitions. It is also possible to carry out this kind of phenotypic fractionation within a QTL-based framework; for example studies have looked directly for correlations between genotypes and quantitative test scores tapping each hypothesized phenotypic component (phoneme awareness, phonological decoding, etc.) in families with dyslexia (Fisher et al., 1999; Gayán et al., 1999). QTL methods allow all phenotyped individuals from a pedigree to be incorporated into the analysis regardless of affection status, but the initial ascertainment of families still relies on diagnostic criteria (Fisher & DeFries, 2002).

What have we learned about the genetic basis of dyslexia by fractionating the overall phenotype into hypothetical components? In their pioneering study, Grigorenko and colleagues (1997) analysed six extended families under different categorical classification schemes representing phenotypes of phoneme awareness, phonological decoding, rapid automised naming, single-word reading, and IQ-reading discrepancy. They targeted two chromosomal regions for investigation, one on chromosome 6 and the other on chromosome 15, because these had been suggested by previous genetic studies (Cardon et al., 1994; Smith, Kimberling, Pennington, & Lubs, 1983), and assessed linkage to each phenotype in their families. Intriguingly the pattern of results with respect to phenotype definitions differed for the two loci; chromosome 6 gave strongest evidence of linkage to impairments in phoneme awareness, while chromosome 15 results were most significant for deficits in single-word reading. Such a dissociation of genetic influences might be taken to imply existence of a "phoneme awareness gene" on chromosome 6 and a "single-word reading gene" on chromosome 15, suggesting that we will ultimately be able to dissect the overall dyslexia phenotype into component-specific genetic effects each localised to a differ-

ent genomic region. So does this mean that there are straightforward pathways between specific genes and certain cognitive processes, after all?

In attempting to answer this question, various authors have raised methodological issues regarding the apparent dissociation of genetic effects observed by Grigorenko et al. (1997); these are of a technical nature and the interested reader is referred to other publications (Field & Kaplan, 1998; Fisher et al., 1999; Fisher & DeFries, 2002; Pennington, 1997). However, beyond the technical caveats that are peculiar to that study, any genetic analysis of dissected phenotypes encounters general constraints that must limit the conclusions we are able to draw. The main problem is that genotype-phenotype correlations can be influenced by many factors that are unrelated to the magnitude of the underlying genetic effects (Fisher & DeFries, 2002; Marlow et al., 2003). These other factors include the sensitivity of diagnostic instruments, the shape of the distribution of test scores, the ages of affected/unaffected individuals and fluctuations in sample size under different diagnostic schemes. Thus, when the same set of families (or individuals) is studied using multiple different, but related, phenotypes it is difficult to be sure that changes in the apparent significance of genotype-phenotype correlation truly reflect aetiological differences at the molecular level. This applies regardless of whether phenotypes are defined in a categorical manner or as fully variable quantitative traits.

The bottom line is that the sample sizes of studies to-date (typically involving hundreds of subjects) may be sufficient for detection of complex genetic effects, but they lack adequate power for determining the relative effect sizes of genes on different measures or phenotypes. As a consequence the profile of results associated with linkage at a particular chromosomal locus can shift substantially with the addition of more families/subjects to the sample. Notably, expansion of the original Grigorenko et al. (1997) sample led to a reversal of the original pattern of results on chromosome 6, such that single-word reading yielded the most significant linkage while phoneme awareness became the weakest-linked phenotype (Grigorenko, Wood, Meyer, & Pauls, 2000), which is of course incompatible with the view that this locus harbours a gene that specifically impacts on phoneme awareness. In addition, studies of other sets of families have provided support for linkage or association on chromosomes 6 and 15, but do not find component-specific effects for either putative risk locus (Deffenbacher et al., 2004; Fisher et al., 1999, 2002b; Francks et al., 2004; Gayán et al., 1999; Kaplan et al., 2002; Schulte-Körne et al., 1998; Turic et al., 2003). For example independent QTL-based studies have replicated linkage to the chromosome 6 locus, but implicated multiple reading- and language-related measures, including phoneme awareness, phonological decoding and orthographic processing (Fisher et al., 1999; Gayán et al., 1999). Similarly, association analyses of this locus point to effects that impact on multiple measures (Deffenbacher et al., 2004; Francks et al., 2004; Kaplan et al., 2002; Turic et al., 2003).

The mapping of a putative susceptibility locus on chromosome 18 provides another compelling example of the potential hazards involved when interpreting data from multiple phenotypes. In a QTL-based genome-wide scan of dyslexia families from the UK, the strongest evidence for linkage was found in a region on the short arm of chromosome 18 (Fisher et al., 2002b). At this locus, results were highly significant

for single-word reading, but much weaker for the other measures employed in the study, including spelling, phoneme awareness, phonological decoding and orthographic processing. Thus, at first glance it might have been tempting to speculate that this chromosome-18 locus harbours a gene with specific effects on single-word reading. In an independent sample of UK families, strong linkage was again seen, to the same region of chromosome 18. However, in this replication sample results were highly significant for the measure of phoneme awareness and weaker for the other measures, with least convincing results for single-word reading. Such a finding argues against specificity of the locus, but, more fundamentally, it raises the question of whether the results from this second sample do in fact constitute replication of the linkage observed in the first sample. These issues were explored in a follow-up study by Marlow and colleagues (2003) who developed a multivariate approach which was able to consider linkage to multiple measures simultaneously. Bivariate and multivariate quantitative analyses of twins had previously proved to be a valuable tool for estimating the overall extent to which different aspects of cognitive ability may share common genetic (and environmental) influences (Gayán & Olson, 2001; Hohnen & Stevenson, 1999). Marlow et al. (2003) extended standard multivariate approaches to facilitate linkage testing of genomic loci in samples of sibling pairs; the covariance between correlated measures is directly incorporated into the linkage analysis, which can yield power gains over comparable univariate methods. Crucially, multivariate linkage methods allow formal assessments of the effect for each trait, considered simultaneously in the context of all other measures, and thus aid in defining which traits contribute to the linkage or replication of a region. Marlow et al. (2003) reanalysed the data from Fisher et al. (2002b) in a multivariate framework and found that (i) the chromosome 18 locus was indeed strongly supported by both the first and second datasets, and (ii) in each case multiple measures were contributing to linkage.

As mentioned above, as well as chromosomes 6, 15 and 18, mapping studies have implicated potential risk loci influencing dyslexia in several other genomic regions (e.g. de Kovel et al., 2004; Fagerheim et al., 1999; Fisher et al., 2002b; Nopola-Hemmi et al., 2001; Kaminen et al., 2003; Tzenova, Kaplan, Petryshen, & Field, 2004). Critically, as yet no study has identified a locus with a robust component-specific influence. This is also the case for investigations of SLI; studies of families with SLI have identified significant linkage of a region on chromosome 16 to repetition of nonsense words, but not to standardised scales assessing overall language (expressive or receptive), which might suggest a component-specific effect (SLI Consortium, 2002). However linkage was also observed at this same locus for measures of reading and spelling in the same sets of families, supporting a more complicated picture (SLI Consortium, 2004). Of course this does not mean that pure component-specific effects will never be discovered, although my guess is that this is unlikely in light of what we already know about molecular genetic systems and the developmental biology of the brain. I should emphasise that I am not arguing here for some kind of genetically encoded "g-factor" for reading and/or language. Even in the absence of straightforward component-specific effects, it is reasonable to expect that allelic variation in different genes may contribute *differentially* to variability in alternative

cognitive measures (e.g. a hypothetical risk allele could perhaps have a greater impact on phoneme awareness than on phonological coding). However, the clear message from studies of dyslexia is that comparison of univariate linkage results for multiple related traits is inadequate for assessing the true contributions of a locus to different aspects of a disorder. Multivariate linkage methods go some way towards addressing this, but even with the availability of the latter, accurate quantification of effect sizes remains something of a challenge. We may need to identify the functional risk variants at each locus and recruit substantially greater numbers of families (perhaps thousands) before we can properly answer these kinds of questions.

Before leaving this discussion, it is worth stressing again that while the measures used in studies of dyslexia may hopefully take us closer to the biological basis of disorder, they are not necessarily pure indices of specific cognitive processes or the activities of individual sets of neural circuits. As noted by Goldberg and Weinberger (2004), the way in which we parse cognitive processes for our phenotypic studies will seldom accurately reflect the underlying genetic architecture. This represents a further level of complexity, over and above molecular and developmental constraints, that should be taken into account when formulating theories based on genetic dissection.

6. The genetics of language: Debunking the "grammar gene" myth

Perhaps the best illustration of the pitfalls associated with an abstract view of the gene revolves around the controversy of so-called "grammar genes" – genes that are specific to aspects of human grammatical ability. Most debates regarding the supposed existence of such genes have been conducted in fields outside molecular genetics or developmental biology and have thus made little or no reference to the inherent limitations and complexities of molecular systems. Research in the area of linguistics first led to the proposal of an innate "universal grammar" constraining the structures of human languages (Chomsky, 1980). Defining the molecular and/or neural bases of this has since become something of a holy grail for the cognitive sciences (Marcus, Vouloumanos, & Sag, 2003). In the early 1990's there was widely publicised media coverage regarding one large three-generational family (known as KE) in which many members suffered from a purportedly grammar-specific disorder, apparently explained by a single gene. This was predominantly based on a short correspondence published in Nature (Gopnik, 1990), which reported that affected family members were unable to infer general rules about signifiers of grammatical features (a deficit referred to as "feature blindness"). Gopnik commented that these deficits "show up in spontaneous speech, writing, grammatical judgement and repetition", suggesting that "the roots probably lie in the underlying grammar" and further noted that the inheritance pattern indicated involvement of just one gene. It is not surprising that this brief report received such a large amount of attention; it was perceived by some as the first hard evidence that human beings possess genes that specify particular aspects of grammar, substantiation of a solid biological basis for linguistic concepts of universal grammar. Despite the fact that the gene causing

the disorder of this family had not even been identified, media coverage hailed this as the discovery of a ''grammar gene'' (for further details see Pinker, 1994).

Over a decade later, we now know the identity of the gene that is damaged in affected members of this family, and are beginning to piece together what it does at molecular and developmental levels (Lai et al., 2001; Lai, Gerrelli, Monaco, Fisher, & Copp, 2003). Thus, for the first time, we are in a good position to connect hypotheses originating in linguistics with data based on molecular/developmental genetics. There is no doubt that the gene (known as *FOXP2*) is relevant to linguistic ability. However, as I will illustrate below, any characterisation of this as a ''gene for grammar'' (or even as a ''gene for language'') clearly becomes untenable once we are able to view it within a more complete biological framework. Furthermore, we also now understand much more about the nature of the disorder observed in the KE family. A series of detailed studies have indicated that deficits in grammatical ability of affected family members are just one aspect of a multifaceted disorder of speech and language (Gopnik & Crago, 1991; Ullman & Gopnik, 1999; Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995; Vargha-Khadem et al., 1998; Watkins, Dronkers, & Vargha-Khadem, 2002). A profound feature of the disorder is a difficulty in coordinating the complex sequences of mouth movements underlying speech, and this is accompanied by wide-ranging linguistic and grammatical impairment. In reactions to the popular myth that the KE family disorder is grammar-specific, it has sometimes been suggested that the disorder is either (i) simply a generalised movement disorder with only secondary effects on language/grammar or (ii) a form of mental retardation affecting both verbal and nonverbal domains. However, these perspectives of the phenotype underplay the importance of linguistic/grammatical features to an inappropriate extent; the true picture lies somewhere in the middle of these extreme views (Marcus & Fisher, 2003). The disorder associated with *FOXP2* disruption affects speech, language and grammar; neither general cognitive deficits nor a basic motor deficit can convincingly account for the profile of impairments, and neuroimaging indicates unambiguous functional abnormalities in neural circuitry related to language processing (Liegeois et al., 2003). I will not dwell further on the issue of phenotypic characterisation here, for more comprehensive discussion the reader is referred to reviews by Fisher et al. (2003), Marcus and Fisher (2003) and Vargha-Khadem, Gadian, Copp, and Mishkin (2005).

Despite the multifaceted nature of the phenotype, the inheritance of disorder in the KE family suggested involvement of only a single gene, facilitating the use of a standard positional strategy to track down its identity. In 2001, Lai and colleagues reported that a change to a single nucleotide in one copy of the *FOXP2* gene on chromosome 7 was responsible for the speech and language problems of the affected KE individuals. They also identified an unrelated subject with speech and language impairment resulting from a translocation disrupting the *FOXP2* locus (Lai et al., 2000; Lai et al., 2001). Although at the time of its discovery *FOXP2* was a novel human gene (in that nobody had previously reported its full coding sequence) it was possible to make predictions about the gene's likely function by comparing it to other genes that had already been characterised. This comparison revealed that *FOXP2* codes for a type of regulatory protein, called a transcription factor, which

is involved in modulating expression of (i.e. switching on and off) other genes (Lai et al., 2001). As I explained earlier in this article, genomic biology is characterised by regulatory networks involving control regions in genes and regulatory factors that bind to them. This feature allows a static linear genome to encode the dynamic adaptive systems underlying the development and functions of a complex living organism (Hood & Galas, 2003). The FOXP2 protein belongs to a subclass of transcription factors known as forkhead proteins (Carlsson & Mahlapuu, 2002), each of which contains a DNA-binding domain (called a ''forkhead-box'' domain) with a characteristic structure. The human genome codes for more than 40 different types of forkhead protein, and these are involved in a wide variety of developmental and physiological pathways (Carlsson & Mahlapuu, 2002). Of note, many forkheads play important roles in controlling genetic cascades during embryonic development and a number are critical for normal patterning of the central nervous system (CNS).

So how does the discovery of *FOXP2* help to account for the disordered speech and language abilities observed in the KE family? Affected KE individuals have a mutation that changes an amino acid at one point of the DNA-binding domain in the FOXP2 protein (Lai et al., 2001). This alteration probably interferes with the function of the mutated protein, so that it is not able to regulate its target genes properly. Note that the disorder in the KE family is dominant, i.e. affected members have one mutant copy of *FOXP2*, and one normal copy. Thus, the speech and language problems appear to be associated with *reduced* levels of functional FOXP2 protein in the brain (rather than a total absence of FOXP2 altogether). Recently, MacDermot et al. (2005) identified a different point mutation in multiple affected members of a newly identified family with speech and language disorder. In this case, the mutation produces a ''stop'' signal in the middle of the FOXP2 protein, severely truncating it, and presumably leading to loss of function. Again these individuals carry one normal copy of *FOXP2* and one copy that is mutated. Similar observations come from studies of chromosomal rearrangements involving *FOXP2*; disruption of only one copy of *FOXP2* is enough to lead to speech and language impairment. Perturbed development due to reduced gene dosage (called ''haploinsufficiency'') is often seen with other developmental disorders associated with transcription factor disruption (Lehmann, Sowden, Carlsson, Jordan, & Bhattacharya, 2003). For example, a range of different mutations in a forkhead gene called *FOXC1* cause disorders of eye development, via the same kind of haploinsufficiency mechanism (Saleem, Banerjee-Basu, Berry, Baxevanis, & Walter, 2003). Patients with *FOXC1* mutations still have one intact copy of *FOXC1*, but the amount of functional FOXC1 protein made by that is not enough for normal eye development.

It is reasonable to conclude from the above studies that reduced amounts of functional FOXP2 protein can lead to disordered brain development or function, in a manner that primarily interferes with speech and/or language abilities. This may well give us clues to the usual function of *FOXP2* in the normal CNS, and perhaps lead to new insights into molecular pathways involved in speech and language acquisition. However, it is essential to realise that this is emphatically *not the same* as saying that *FOXP2* is a ''gene for speech'' or a ''gene for language'', that its raison d'etre is to provide humans with speech/language abilities. This is not merely a question of

semantics, since terminology can have a profound impact on conceptual viewpoints, as I expand on at the end of this article. The invalidity of the "language gene" / "speech gene" tag becomes even more apparent as we consider additional features of the gene that have been uncovered. First, *FOXP2* mutation (although not restricted to the KE family) accounts for only a small proportion of cases of disordered speech and language (MacDermot et al., 2005). Second, *FOXP2* is not unique to humans, and is found in extremely similar form throughout mammalian species (Enard et al., 2002; Zhang, Webb, & Podlaha, 2002). The protein encoded by the mouse version of this gene differs from the human protein at only three amino acid positions out of more than 700. Since mice clearly lack linguistic capabilities, the remarkably high conservation of *FOXP2* across the species is incompatible with the idea that the gene exists solely to subserve the capacity for speech and language. In other words, we can conclude that functional FOXP2 protein was already present in the common nonlinguistic ancestor of humans and rodents over 70 million years ago in a very similar form to that found in modern humans. Third, studies examining the expression of *FOXP2* (i.e. when and where it is switched on during development and in adulthood) indicate that its functional importance is not restricted to the CNS (Shu, Yang, Zhang, Lu, & Morrisey, 2001). The way that the patterns of gene expression unfold in the developing embryo suggests that *FOXP2* regulates key pathways in the developing lung, heart and gut. For example, in the developing lung, expression of the gene becomes restricted to the distal airway epithelium, and experiments have shown that it is able to regulate other genes that are important for lung function (Shu et al., 2001). The recycled use of the same regulatory factors to control multiple pathways in different developmental contexts is a common feature of complex biological systems; it is rare to find a transcription factor that has an exclusive role specific to only one context. Thus, calling *FOXP2* a "language gene" makes no more sense than referring to it as a "lung gene" or a "distal airway epithelium gene". Fourth, detailed studies of *FOXP2* in the brain indicate complexities of expression even when we limit our examination just to the CNS. In both humans and rodents, the gene is switched on in a range of brain regions during early development, including cortex, thalamus, hypothalamus, striatum, cerebellum and medulla, and expression persists into adulthood (Ferland, Cherry, Preware, Morrisey, & Walsh, 2003; Lai et al., 2003; Takahashi, Liu, Hirokawa, & Takahashi, 2003). A naïve view of a gene that exists to provide us with linguistic capabilities would predict an expression pattern that is restricted to a region of the brain such as Broca's area.

In sum, much of the data on *FOXP2* from molecular and developmental biology confounds any expectations that one might have for a hypothetical "language gene"; and the reason for this is that this entire concept is flawed, being rooted in an abstract view of the nature of the gene. *FOXP2* is a well conserved regulatory gene with multiple roles during development (and perhaps also in adulthood) and is likely to influence patterning/function of several regions of the CNS in all mammals (Lai et al., 2003). Nevertheless it is essential to emphasise that the above considerations do not diminish the significance of this gene for understanding molecular pathways underlying speech and language (Marcus & Fisher, 2003). Rather, they provide a case-study showing that biological systems do not conform to the view of a simple relationship between genes and

cognitive development, even when it comes to single-gene disorders. Importantly, this does not mean that the systems are impenetrable; instead we can begin to incorporate molecular and developmental complexity into our models of what *FOXP2* does. For example, although the gene is indeed switched on in several areas of the developing and mature CNS, detailed evaluation indicates that the expression pattern is tightly regulated in each region. Expression is highest in the deep layers in the cortex, the striosomal compartment in the striatum, the Purkinje cells and deep nuclei in the cerebellum and the inferior olives in the medulla, and the timing of onset suggests a role in neuronal development that is post-proliferation/migration in these regions (Ferland et al., 2003; Lai et al., 2003; Takahashi et al., 2003). These findings are compatible with the hypothesis that *FOXP2* helps to establish and maintain connectivity of corticostriatal and olivocerebellar neural circuitry. (This idea is currently speculative, but is one that can actually be tested in future in animal models using standard approaches of molecular neuroscience.) The concordance between rodent and human expression patterns suggests that *FOXP2* was already playing such a role in the common ancestor of these species (Lai et al., 2003).

Despite the high conservation of *FOXP2* coding sequence among mammals, and the strong likelihood that the gene was involved in patterning corticostriatal/olivocerebellar circuitry many millions of years before the emergence of speech and language, there is intriguing evidence that modifications of the gene may have been selected in recent human evolution (Enard et al., 2002; Zhang et al., 2002). Of the three amino acid changes that distinguish the mouse and human versions of the FOXP2 protein, two occurred on the human lineage after the split between humans and chimps, and statistical analyses suggest that these arose and rapidly spread through the human population at some point within the last 200,000 years. The timing of this is concordant with estimates for the emergence of human speech and language (around 50,000–100,000 years ago) (Boyd & Silk, 2000). These data have led to something of a resurgence of the ''speech gene''/''language gene'' tag in the media, a common theme being that tiny substitutions created radical changes in FOXP2 function and were the *major driving force* behind the evolution of human speech and language. Again, consideration of the molecular and developmental context argues against this extreme viewpoint. At present there is nothing known about the functional significance of the two amino acid changes that occurred on the human lineage; they lie outside the well-characterised domains of the protein. The inference of functional significance is primarily based on statistical analyses of sequence diversity. One of the changes has the potential to create a new site for regulation of the protein by other proteins; however, no-one has yet tested this hypothesis directly, and in fact this same change occurred independently in carnivores (Zhang et al., 2002), which is difficult to reconcile with the idea that its appearance in humans was the driving force for speech and language evolution. It is also worth remembering that the human-specific version of *FOXP2* is still essential for development of other tissues outside of the brain (Lai et al., 2001), so any modifications should have preserved the ''original'' functions of the gene to a large extent.

Integration of all the *FOXP2* findings supports a nonsensational, but still tantalizing, picture of the recent evolution of this gene. It is reasonable to assume that *FOXP2*

was already involved in patterning corticostriatal and olivocerebellar circuitry in the CNS of our nonlinguistic ancestors. It has long been known that these distributed circuits are critical for complex motor control. There is an emerging consensus in the field of neuroscience that they are also key for aspects of cognition, such as procedural learning, defined as the unconscious acquisition of skills through practice (Middleton & Strick, 2000; Welsh, Lang, Suglhara, & Llinas, 1995). At a time when vocal communication was emerging in humans (for other undetermined reasons, be they genetic and/ or environmental), changes in *FOXP2* which improved motor sequencing and/or procedural learning abilities (via modifications to corticostriatal/olivocerebellar patterning) might have given those humans possessing them a selective advantage. Thus, given an environment in which vocal communication was becoming important, the changes would have rapidly spread through the population. In such an environmental context, a significant selective advantage may have resulted from only subtle tweaking of existing pathways (e.g. increased connectivity of corticostriatal circuits). Remember that radical changes to FOXP2 would be unlikely to preserve its critical role in other tissues (Lai et al., 2001; Shu et al., 2001), and thus might be lethal in the homozygous state (i.e. when both copies of *FOXP2* are altered). Of course, we will probably never really know the true course of events that gave rise to fixation of the modern human version of FOXP2; the reasons for selection may not have related to speech and language at all (Lieberman, 2002). However, to my mind, the model I have described here is the most parsimonious based on current knowledge.

There remain many open questions regarding the role of the *FOXP2* gene, which lie beyond the scope of the present article. The most prescient point of this story is that researchers now have the means to directly address many of the relevant issues via molecular and developmental studies, and can attempt to formally integrate any findings with those from investigations at other levels (e.g. neuroimaging, psycholinguistics etc.). For example, there are intriguing concordances between the regions of expression of *FOXP2* (Lai et al., 2003) and adult sites of brain pathology revealed by structural and functional neuroimaging (Liegeois et al., 2003; Vargha-Khadem et al., 1998). Moreover, both approaches support the suggestion from the field of cognitive sciences that the KE family disorder involves deficits in procedural learning (Ullman & Gopnik, 1999), although it is still unclear whether such deficits explain all the linguistic problems of affected individuals. Finally, the complexity of expression patterns may help to account for the fact that *FOXP2* disruption leads to impairment in multiple aspects of CNS function (involving multiple brain regions). This suggests an interesting possibility; even though we know the primary genetic cause of the disorder, there may not in fact be a single identifiable core deficit at the neurological level.

7. The language of genetics: Concluding remarks

Looking to the future, it is apparent that the integrated efforts of multiple disciplines hold great promise for mapping the connections between genes and cognition (Grigorenko, 2001; Hauser, Chomsky, & Fitch, 2002). For this to be fruitful, it is important that scientists who are already experts in their own fields acquire at least a basic fluency

in the language of other research areas. Geneticists need to become familiar with the limitations of diagnosing neurodevelopmental disorders, the ways that cognitive phenotypes are parsed at the neuropsychological level, and the conclusions that may be drawn from neuroimaging studies; linguists need to understand the routes by which genetic effects can be mapped and appreciate the restrictions imposed by molecular systems; and so on. This will require some extra effort, but all may then reap the benefits offered by a truly holistic view of neural development and function.

The common feature of the examples that I have presented in the current article is a misconception of simple linear relations between genes and behavioural or cognitive outputs. The misconception is fuelled in part by a language barrier between the different fields. Geneticists routinely use the "gene for X" phraseology as a convenient shorthand to describe their results. It is easier to refer to the "gene for cystic fibrosis" than "the chloride channel gene which when it is mutated gives rise to cystic fibrosis". Similarly, when geneticists talk about an "asthma gene" in a common population they really mean "a gene, allelic variants of which influence a person's risk of developing asthma". It is clear that this kind of terminology has considerable value in that it simplifies communication within the field, and it is important for increasing public awareness regarding genetic contributions to traits, where references to complexity can become ineffective. Nevertheless, I suggest that the common use of phrases like "gene for dyslexia" and "language gene" has helped many of those outside of molecular fields to come to erroneous conclusions. I am not proposing that such phrases have no value and should be outlawed, but for those researchers seeking to unite genetics and cognition, I urge the following. First, geneticists should aim for increased clarity when reporting their results in order to avoid misinterpretation by others, and second, non-geneticists should aim to become familiar with the language of genetics, including all the caveats that accompany the phrase "a gene for X".

I have implied here that many scientists investigating cognition are currently happy to view genes as abstract entities, without much concern for the biological mechanisms by which genes act to influence brain development or function. It could be argued that this is an unfair criticism, since, in general, the functional genetic variants influencing common forms of neurodevelopmental disorder are unknown. Perhaps researchers do become interested in molecular systems and developmental constraints, but only once the relevant genes themselves are clearly known. This argument seems persuasive, but misses the point somewhat. Of course it is unreasonable to expect scientists to grasp the molecular minutiae underlying reading dysfunction (say) prior to confirmation of the key genetic players. However it is important to distinguish this kind of specific mechanistic understanding from a broader appreciation of the general ways that genes are able to function to help build and maintain a working brain. A basic knowledge of what genes can and cannot do is necessary to constrain and inform hypotheses about how genetic variability may relate to cognitive variability (whether in the normal range or with respect to disorder). Crucially, such hypotheses can actually play a fundamental role in our gene searches. For example, if researchers had adopted a human-specific "language gene" view when searching for the molecular basis of disorder in the KE family, then they may well have discounted *FOXP2* at

an early stage, since the gene is so highly conserved in other (nonlinguistic) species. Moreover, it is not obvious that cognitive scientists *do* become interested in the molecular pathways once the relevant genes are identified; even now discussions regarding the role of *FOXP2* in human speech and language do not tend to exploit the relevant emerging data from molecular studies, but remain focused on approaches that were possible prior to isolation of the gene (Bosman, Garcia, & Aboitiz, 2004; Corballis, 2004).

Putting all these issues aside, it remains to be seen whether or not we will ever identify a gene which appears to have specific influences on particular aspects of cognition, such as phoneme awareness, or grammatical processing, but I doubt that nature will present us with anything quite so tidy. Consideration of everything we have learned thus far from genetic studies of neurodevelopmental disorders leads to one inescapable conclusion. If we are able to place findings into a framework of systems biology that acknowledges molecular and developmental complexity, we will greatly improve our chances of untangling the webs that link genes to cognition.

Acknowledgements

References

Arcos-Burgos, M., Castellanos, F. X., Pineda, D., Lopera, F., Palacio, J. D., Palacio, L. G., et al. (2004). Attention-deficit/hyperactivity disorder in a population isolate: linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. *American Journal of Human Genetics, 75*, 998–1014.

Bakker, S. C., van der Meulen, E. M., Buitelaar, J. K., Sandkuijl, L. A., Pauls, D. L., Monsuur, A. J., et al. (2003). A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *American Journal of Human Genetics, 72*, 1251–1260.

Bartlett, C. W., Flax, J. F., Logue, M. W., Vieland, V. J., Bassett, A. S., Tallal, P., et al. (2002). A major susceptibility locus for specific language impairment is located on 13q21. *American Journal of Human Genetics, 71*, 45–55.

Bell, J. I. (2003). The double helix in clinical practice. *Nature, 421*, 414–416.

Bosman, C., Garcia, R., & Aboitiz, F. (2004). FOXP2 and the language working-memory system. *Trends in Cognitive Sciences, 8*, 251–252.

Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics, 33*(Suppl), 228–237.

Boyd, R., & Silk, J. B. (2000). *How humans evolved*. New York: W.W. Norton.

Buxbaum, J. D., Silverman, J. M., Smith, C. J., Kilifarski, M., Reichert, J., Hollander, E., et al. (2001). Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *American Journal of Human Genetics, 68*, 1514–1520.

Carlsson, P., & Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. *Developmental Biology, 250*, 1–23.

Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics, 2*, 91–99.

Cardon, L. R., Smith, S. D., Fulker, D. W., Kimberling, W. J., Pennington, B. F., & DeFries, J. C. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science, 266*, 276–279.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149–180.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Collaborative Linkage Study of Autism (CLSA) (2001). An autosomal genomic screen for autism. *American Journal of Medical Genetics, 105*, 609-615.

Collins, F. S., Green, E. D., Guttmacher, A. E., & Guyer, M. S. (2003). A vision for the future of genomics research. *Nature, 422*, 835–847.

Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., et al. (2005a). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *American Journal of Human Genetics, 76*, 581–591.

Cope, N. A., Hill, G., vandenBree, M., Harold, D., Moskvina, V., Green, E. K., et al. (2005b). No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. *Molecular Psychiatry, 10*, 237–238.

Corballis, M. C. (2004). FOXP2 and the mirror system. *Trends in Cognitive Sciences, 8*, 95–96.

Deffenbacher, K. E., Kenyon, J. B., Hoover, D. M., Olson, R. K., Pennington, B. F., DeFries, J. C., et al. (2004). Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Human Genetics, 115*, 128–138.

de Kovel, C. G., Hol, F. A., Heister, J. G., Willemen, J. J., Sandkuijl, L. A., Franke, B., et al. (2004). Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *Journal of Medical Genetics, 41*, 652–657.

DiMaio, S., Grizenko, N., & Joober, R. (2003). Dopamine genes and attention-deficit hyperactivity disorder: a review. *Journal of Psychiatry and Neuroscience, 28*, 27–38.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature, 418*, 869–872.

Fagerheim, T., Raeymaekers, P., Tonnessen, F. E., Pedersen, M., Tranebjaerg, L., & Lubs, H. A. (1999). A new gene (DYX3) for dyslexia is located on chromosome 2. *Journal of Medical Genetics, 36*, 664–669.

Ferland, R. J., Cherry, T. J., Preware, P. O., Morrisey, E. E., & Walsh, C. A. (2003). Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *Journal of Comparative Neurology, 460*, 266–279.

Field, L. L., & Kaplan, B. J. (1998). Absence of linkage of phonological coding dyslexia to chromosome 6p23-p21.3 in a large family data set. *American Journal of Human Genetics, 63*, 1448–1456.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nature Reviews Neuroscience, 3*, 767–780.

Fisher, S. E., Francks, C. (2006). Genes, cognition and dyslexia: learning to read the genome. *Trends in Cognitive Sciences*, 10, in press.

Fisher, S. E., Francks, C., McCracken, J. T., McGough, J. J., Marlow, A. J., MacPhie, I. L., et al. (2002a). A genomewide scan for loci involved in attention-deficit/hyperactivity disorder. *American Journal of Human Genetics, 70*, 1183–1196.

Fisher, S. E., Francks, C., Marlow, A. J., MacPhie, I. L., Newbury, D. F., Cardon, L. R., et al. (2002b). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature Genetics, 30*, 86–91.

Fisher, S. E., Lai, C. S. L., & Monaco, A. P. (2003). Deciphering the genetic basis of speech and language disorders. *Annual Review of Neuroscience, 26*, 57–80.

Fisher, S. E., Marlow, A. J., Lamb, J., Maestrini, E., Williams, D. F., Richardson, A. J., et al. (1999). A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *American Journal of Human Genetics, 64*, 146–156.

Fisher, S. E., & Marcus, G. F. (2006). The eloquent ape: genes, brains and the evolution of language. *Nature Reviews Genetics, 7*, 9–20.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics, 18*, 168–170.

Flint, J., & Knight, S. (2003). The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation. *Current Opinion in Genetics and Development, 13*, 310–316.

Folstein, S. E., & Rosen-Sheidley, B. (2001). Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Reviews Genetics, 2*, 943–955.

Francks, C., Fisher, S. E., Olson, R. K., Pennington, B. F., Smith, S. D., DeFries, J. C., et al. (2002). Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes SEMA4F and OTX1. *Psychiatric Genetics, 12*, 35–41.

Francks, C., Paracchini, S., Smith, S. D., Richardson, A. J., Scerri, T. S., Cardon, L. R., et al. (2004). A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *American Journal of Human Genetics, 75*, 1046–1058.

Freimer, N., & Sabatti, C. (2003). The human phenome project. *Nature Genetics, 34*, 15–21.

Gauthier, J., Bonnel, A., St-Onge, J., Karemera, L., Laurent, S., Mottron, L., et al. (2004). NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *American Journal of Medical Genetics Part B, 132*, 74–75.

Gayán, J., & Olson, R. K. (2001). Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. *Developmental Neuropsychology, 20*, 483–507.

Gayán, J., Smith, S. D., Cherny, S. S., Cardon, L. R., Fulker, D. W., Brower, A. M., et al. (1999). Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *American Journal of Human Genetics, 64*, 157–164.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science, 298*, 2345–2349.

Goldberg, T. E., & Weinberger, D. R. (2004). Genes and the parsing of cognitive processes. *Trends in Cognitive Sciences, 8*, 325–335.

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature, 344*, 715, [correspondence].

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition, 39*, 1–50.

Grant, S. G. (2003). Systems biology in neuroscience: bridging genes to cognition. *Current Opinion in Neurobiology., 13*, 577–582.

Grigorenko, E. L. (2001). Developmental dyslexia: an update on genes, brains, and environments. *Journal of Child Psychology and Psychiatry, 42*, 91–125.

Grigorenko, E. L., Wood, F. B., Meyer, M. S., Hart, L. A., Speed, W. C., Shuster, A., et al. (1997). Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *American Journal of Human Genetics, 60*, 27–39.

Grigorenko, E. L., Wood, F. B., Meyer, M. S., & Pauls, D. L. (2000). Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *American Journal of Human Genetics, 66*, 715–723.

Habib, M. (2000). The neurological basis of developmental dyslexia: an overview and working hypothesis. *Brain, 123*, 2373–2399.

Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kaariainen, H., et al. (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *Public Library of Science Genetics, 1*, e50.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve?. *Science 298*, 1569–1579.

Hohnen, B., & Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology, 35*, 590–603.

Hood, L., & Galas, D. (2003). The digital code of DNA. *Nature, 421*, 444–448.

Inoue, K., & Lupski, J. R. (2003). Genetics and genomics of behavioral and psychiatric disorders. *Current Opinion in Genetics and Development, 13*, 303–309.

International Molecular Genetic Study of Autism Consortium (IMGSAC). (1998). A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Human Molecular Genetics*, 7, 571–578.

International Molecular Genetic Study of Autism Consortium (IMGSAC) (2001). A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *American Journal of Human Genetics*, 69, 570–581.

Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I. C. et al. Paris Autism Research International Sibpair Study. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics*, 34, 27–29.

Kakiuchi, C., Iwamoto, K., Ishiwata, M., Bundo, M., Kasahara, T., Kusumi, I., et al. (2003). Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. *Nature Genetics, 35*, 171–175.

Kaminen, N., Hannula-Jouppi, K., Kestila, M., Lahermo, P., Muller, K., Kaaranen, M., et al. (2003). A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *Journal of Medical Genetics, 40*, 340–345.

Kaplan, D. E., Gayaín, J., Ahn, J., Won, T. W., Pauls, D., Olson, R. K., et al. (2002). Evidence for linkage and association with reading disability on 6p21.3-22. *American Journal of Human Genetics, 70*, 1287–1298.

Korstanje, R., & Paigen, B. (2002). From QTL to gene: the harvest begins. *Nature Genetics, 31*, 235–236.

Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., & Gage, F. H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell, 116*, 779–793.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Levy, E. R., Hodgson, S., Fox, M., et al. (2000). The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *American Journal of Human Genetics, 67*, 357–368.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A novel forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*, 519–523.

Lai, C. S. L., Gerrelli, D., Monaco, A. P., Fisher, S. E., & Copp, A. J. (2003). FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain, 126*, 2455–2462.

Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics, 11*, 241–247.

Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science, 265*, 2037–2048.

Lehmann, O. J., Sowden, J. C., Carlsson, P., Jordan, T., & Bhattacharya, S. S. (2003). Fox's in development and disease. *Trends in Genetics, 19*, 339–344.

Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology, 45*, 36–62.

Liegeois, F., Baldeweg, T., Connelly, A., Gadian, D. G., Mishkin, M., & Vargha-Khadem, F. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nature Neuroscience, 6*, 1230–1237.

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S. L., Vernes, S. C., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American Journal of Human Genetics, 76*, 1074–1080.

Marcus, G. F. (2003). *The birth of the mind: How a tiny number of genes creates the complexities of human thought*. New York: Basic Books.

Marcus, G. F., & Fisher, S. E. (2003). FOXP2 in focus: what can genes tell us about speech and language?. *Trends in Cognitve Sciences 7*, 257–262.

Marcus, G. F., Vouloumanos, A., & Sag, I. A. (2003). Does Broca's play by the rules?. *Nature Neuroscience 6*, 651–652.

Marino, C., Giorda, R., Luisa Lorusso, M., Vanzin, L., Salandi, N., Nobile, M., et al. (2005). A family-based association study does not support DYX1C1 on 15q21.3 as a candidate gene in developmental dyslexia. *European Journal of Human Genetics, 13*, 491–499.

Marlow, A. J., Fisher, S. E., Francks, C., MacPhie, I. L., Cherny, S. S., Richardson, A. J., et al. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *American Journal of Human Genetics, 72*, 561–570.

Marlow, A. J., Fisher, S. E., Richardson, A. J., Francks, C., Talcott, J. B., Monaco, A. P., et al. (2001). Investigation of quantitative measures related to reading disability in a large sample of sib-pairs from the UK. *Behavior Genetics, 31*, 219–230.

Meng, H., Smith, S. D., Hager, K., Held, M., Liu, J., Olson, R. K., et al. (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proceedings of the National Academy of Sciences United States of America, 102*, 17053–17058.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Research Reviews, 31*, 236–250.

Morris, C. A., & Mervis, C. B. (2000). Williams syndrome and related disorders. *Annual Review of Genomics and Human Genetics, 1*, 461–484.

Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S. L., et al., and the International Molecular Genetic Study of Autism Consortium. (2002). FOXP2 is not a major susceptibility gene for autism or Specific Language Impairment (SLI). *American Journal of Human Genetics*, 70, 1318–1327.

Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., et al. (2001). A dominant gene for developmental dyslexia on chromosome 3. *Journal of Medical Genetics, 38*, 658–664.

O'Brien, E. K., Zhang, X., Nishimura, C., Tomblin, J. B., & Murray, J. C. (2003). Association of Specific Language Impairment (SLI) to the Region of 7q31. *American Journal of Human Genetics, 72*, 1536–1543.

Ogdie, M. N., Macphie, I. L., Minassian, S. L., Yang, M., Fisher, S. E., Francks, C., et al. (2003). A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: suggestive linkage on 17p11. *American Journal of Human Genetics, 72*, 1268–1279.

Pääbo, S. (2003). The mosaic that is our genome. *Nature, 421*, 409–412.

Page, G. P., George, V., Go, R. C., Page, P. Z., & Allison, D. B. (2003). Are we there yet?: Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *American Journal of Human Genetics, 73*, 711–719.

Pennington, B. F. (1997). Using genetics to dissect cognition. *American Journal of Human Genetics, 60*, 13–16.

Philippe, A., Martinez, M., Guilloud-Bataille, M., Gillberg, C., Rastam, M., Sponheim, E., et al. (1999). Genome-wide scan for autism susceptibility genes, Paris Autism Research International Sibpair Study. *Human Molecular Genetics, 8*, 805–812.

Pinker, S. (1994). *The language instinct*. London: Allen Lane.

Reich, D. E., & Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends in Genetics, 17*, 502–510.

Roizen, N. J., & Patterson, D. (2003). Down's syndrome. *Lancet, 361*, 1281–1289.

Saleem, R. A., Banerjee-Basu, S., Berry, F. B., Baxevanis, A. D., & Walter, M. A. (2003). Structural and functional analyses of disease-causing missense mutations in the forkhead domain of FOXC1. *Human Molecular Genetics, 12*, 2993–3005.

Scerri, T. S., Fisher, S. E., Francks, C., MacPhie, I. L., Paracchini, S., Richardson, A. J., et al. (2004). Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. *Journal of Medical Genetics, 41*, 853–857.

Schulte-Körne, G., Grimm, T., Nothen, M. M., Muller-Myhsok, B., Cichon, S., Vogt, I. R., et al. (1998). Evidence for linkage of spelling disability to chromosome 15. *American Journal of Human Genetics, 63*, 279–282.

Shao, Y., Raiford, K. L., Wolpert, C. M., Cope, H. A., Ravan, S. A., Ashley-Koch, A. A., et al. (2002). Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *American Journal of Human Genetics, 70*, 1058–1061.

Shu, W., Yang, H., Zhang, L., Lu, M. M., & Morrisey, E. E. (2001). Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *Journal of Biological Chemistry, 276*, 27488–27497.

SLI Consortium (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *American Journal of Human Genetics*, 70, 384–398.

SLI Consortium (2004). Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *American Journal of Human Genetics*, 74, 1225–1238.

Smith, S. D., Kimberling, W. J., Pennington, B. F., & Lubs, H. A. (1983). Specific reading disability: identification of an inherited form through linkage analysis. *Science, 219*, 1345.

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia, 7*, 37–46.

Takahashi, K., Liu, F. C., Hirokawa, K., & Takahashi, H. (2003). Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *Journal of Neuroscience Research, 73*, 61–72.

Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., et al. (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proceedings of the National Academy of Sciences United States of America, 100*, 11553–11558.

Turic, D., Robinson, L., Duke, M., Morris, D. W., Webb, V., Hamshere, M., et al. (2003). Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3-22. *Molecular Psychiatry, 8*, 176–185.

Tzenova, J., Kaplan, B. J., Petryshen, T. L., & Field, L. L. (2004). Confirmation of a dyslexia susceptibility locus on chromosome 1p34-p36 in a set of 100 Canadian families. *American Journal of Medical Genetics Part B, 127*, 117–124.

Ullman, M. T., & Gopnik, M. (1999). Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics, 20*, 51–117.

Vargha-Khadem, F., Gadian, D. G., Copp, A., & Mishkin, M. (2005). FOXP2 and the neuroanatomy of speech and language. *Nature Reviews Neuroscience, 6*, 131–138.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences United States of America, 92*, 930–933.

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., et al. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences United States of America, 95*, 12695–12700.

Vincent, J. B., Kolozsvari, D., Roberts, W. S., Bolton, P. F., Gurling, H. M. D, & Scherer, S. W. (2004). Mutation screening of X-chromosomal neuroligin genes: No mutations in 196 autism probands. *American Journal of Medical Genetics Part B, 129*, 82–84.

Wassink, T. H., Brzustowicz, L. M., Bartlett, C. W., & Szatmari, P. (2004). The search for autism disease genes. *Mental Retardation in Developmental Disabilities Research Reviews, 10*, 272–283.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain, 125*, 452–464.

Welsh, J. P., Lang, E. J., Suglhara, I., & Llinas, R. (1995). Dynamic organization of motor control within the olivocerebellar system. *Nature, 374*, 453–457.

Wigg, K. G., Couto, J. M., Feng, Y., Anderson, B., Cate-Carter, T. D., Macciardi, F., et al. (2004). Support for EKN1 as the susceptibility locus for dyslexia on 15q21. *Molecular Psychiatry, 9*, 1111–1121.

Wolf, M., & Bowers, P. G. (2000). Naming-speed processes and developmental reading disabilities: An introduction to the special issue on the double-deficit hypothesis. *Journal of Learning Disabilities, 33*, 322–324.

Wolfsberg, T. G., McEntyre, J., & Schuler, G. D. (2001). Guide to the draft human genome. *Nature, 409*, 824–826.

Yonan, A. L., Alarcon, M., Cheng, R., Magnusson, P. K., Spence, S. J., Palmer, A. A., et al. (2003). A genomewide screen of 345 families for autism-susceptibility loci. *American Journal of Human Genetics, 73*, 886–897.

Zhang, J., Webb, D. M., & Podlaha, O. (2002). Accelerated Protein Evolution and Origins of Human-Specific Features. Foxp2 as an example. *Genetics, 162*, 1825–1835.